Structure-Guided Design of a Peptide Lock for Modular Peptide Binders

Ernst, Patrick ; Zosel, Franziska ; Reichen, Christian ; Nettels, Daniel ; Schuler, Benjamin ; Plückthun, Andreas

Abstract: Peptides play an important role in intermolecular interactions and are frequent analytes in diagnostic assays, also as unstructured, linear epitopes in whole proteins. Yet, due to the many different sequence possibilities even for short peptides, classical selection of binding proteins from a library, one at a time, is not scalable to proteomes. However, moving away from selection to a rational assembly of preselected modules binding to predefined linear epitopes would split the problem into smaller parts. These modules could then be reassembled in any desired order to bind to, in principle, arbitrary sequences, thereby circumventing any new rounds of selection. Designed Armadillo repeat proteins (dArmRPs) are modular, and they do bind elongated peptides in a modular way. Their consensus sequence carries pockets that prefer arginine and lysine. In our quest to select pockets for all amino acid side chains, we had discovered that repetitive sequences can lead to register shifts and peptide flipping during selections from libraries, hindering the selection of new binding specificities. To solve this problem, we now created an orthogonal binding specificity by a combination of grafting from -catenin, computational design and mutual optimization of the pocket and the bound peptide. We have confirmed the design and the desired interactions by X-ray structure determination. Furthermore, we could confirm the absence of sliding in solution by a single-molecule Förster resonance energy transfer. The new pocket could be moved from the N-terminus of the protein to the middle, retaining its properties, further underlining the modularity of the system.

DOI: https://doi.org/10.1021/acschembio.9b00928

Structure-guided design of a peptide lock for modular peptide binders

Journal:	*ACS Chemical Biology*
Manuscript ID	cb-2019-009289.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Ernst, Patrick; Universitat Zurich, Dept. of Biochemistry Zosel, Franziska; Universitat Zurich, Dept. of Biochemistry Reichen, Christian; Universitat Zurich, Dept. of Biochemistry Nettels, Daniel; Universitat Zurich, Biochemisches Institut Schuler, Benjamin; Universitat Zurich, Biochemisches Institut Plückthun, Andreas; Universitat Zurich, Dept. of Biochemistry

Structure-guided design of a peptide lock for modular peptide binders

Patrick Ernst[1], Franziska Zosel[2], Christian Reichen[3], Daniel Nettels, Benjamin Schuler and Andreas Plückthun[*].

Department of Biochemistry, University Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Corresponding Author

*Andreas Plückthun

Tel. +41-44-635 5570, Fax. +41-44-635 5712

E-mail: plueckthun@bioc.uzh.ch

Present Addresses

[1] Present address: University of Basel, Graduate Center, Petersgraben 35, 4051 Basel, Switzerland

[2] Present address: Novo Nordisk A/S, Novo Nordisk Park 1, 2760 Måløv, Denmark

[3] Present address: Molecular Partners AG, Wagistrasse 14, 8952 Zürich-Schlieren, Switzerland

ABSTRACT:

Peptides play an important role in intermolecular interactions and are frequent analytes in diagnostic assays, also as unstructured, linear epitopes in whole proteins. Yet, due to the many different sequence possibilities even for short peptides, classical selection of binding proteins from a library, one at a time, is not scalable to proteomes. However, moving away from selection to a rational assembly of preselected modules binding to predefined linear epitopes would split the problem into smaller parts. These modules could then be reassembled in any desired order to bind to, in principle, arbitrary sequences, thereby circumventing any new rounds of selection. Designed Armadillo repeat proteins (dArmRPs) are modular and they do bind elongated peptides in a modular way. Their consensus sequence carries pockets that prefer arginine and lysine, respectively. In our quest to select pockets for all amino acid side chains, we had discovered that repetitive sequences can lead to register shifts and peptide flipping during selections from libraries, hindering the selection of new binding specificities. To solve this problem, we now created an orthogonal binding specificity by a combination of grafting from beta-catenin and computational design and mutual optimization of the pocket and the bound peptide. We have confirmed the design and the desired interactions by x-ray structure determination. Furthermore, we could confirm the absence of sliding in solution by single-molecule Förster resonance energy transfer. The new pocket could be moved from the N-terminus of the protein to the middle, retaining its properties, further underlining the modularity of the system.

INTRODUCTION

Protein-peptide interactions are the basis for 15-40% of all interactions in the cell, and many biochemical detection assays rely on the specific interaction of binding molecules or protein domains with unfolded proteins, loops or peptide tags lacking a secondary structure[1]. Furthermore, 35-45% of the eukaryotic proteome is predicted to be unstructured[2]. Thus, linear, unstructured epitopes are of broad interest in protein detection and identification.

Nowadays, many different binding scaffolds can achieve peptide binding; even picomolar affinities have been reported for the interaction between synthetic antibodies and longer peptides[3]. However, the sheer number of potential amino acid sequences, even for short peptides, would make it impossible to stockpile sequence-specific binders for every peptide by classical immunization or classical selection techniques, one binder at a time. For detection of proteins and post-translational modifications on proteomic scales, a rapid access to specific sequence recognition would be transformative. Yet, this problem can be overcome with a binding scaffold in which peptide binding relies solely on the primary sequence. Ideally, such a scaffold is made up of binding modules which recognize short parts of the peptide sequence. In that case, one would only need to generate specific binding pockets or modules for the 20 different amino acids and their modified versions for the detection of post-translational modifications. A reassembly of previously generated modules would then allow generation of binders to, in principle, any given peptide target sequence.

Modular proteins already exist in nature, and the family of natural Armadillo repeat proteins (nArmRPs) binds their target peptides in an elongated conformation and already in a modular manner with, on average, two amino acids per repeat[4–6]. Over the last years, nArmRPs served as the basis for the design of fully modular and highly stable peptide binders, called designed Armadillo repeat proteins (dArmRPs)[7–16] (reviewed in [17,18]). The peptide, which carries alternating arginines and lysines, is bound to specific pockets on the surface of the dArmRP, and the backbone is kept elongated by a conserved ladder of asparagines at position 37 (N37) of each repeat (Figure 1 A). Furthermore, the affinity can be easily adjusted by varying the length of the peptide, or the number of binding modules, respectively[15]. Increasing the number of dipeptide units or binding modules leads to a linear increase of binding energy, with dissociation constants ranging from high nanomolar to low picomolar[15]. Alanine scanning has shown a constant contribution to binding energy from each arginine pocket, and a smaller constant contribution from each lysine pocket.

However, the repetitive nature of both the peptide and the dArmRP can lead to additional, alternative binding modes. Recent crystal structures confirm the existence of those alternative binding modes, in which the peptide binds to the dArmRP in different registers (Figure 1 B)[19]. However, multiple binding modes would be counteracting the selection process, as they prevent the sequence-specific recognition of individual amino acids if alternative residues from a different binding register occupy a given binding pocket. One solution for this problem is the design of new binding specificities, which could lock one binding mode and would thus break the symmetry of the repetitive peptide.

In this paper, we present a design process in which we grafted a hydrophobic binding pocket of β-catenin onto the interaction surface of dArmRPs and describe how computational design was utilized to optimize its affinity and specificity. Computational protein-protein interface design can be a complementary approach for generating new binding proteins, especially with the protein design suite Rosetta[20]. Yet, especially the design of polar interactions can be a hurdle[21]. Nevertheless, many successful examples have been reported[22–25], and pre-defined natural binding interfaces or hot-spot residues may already provide a good starting point for the design[26–28]. For these reasons, we proceeded from a graft to a computational improvement of the pocket.

This pocket prevents the peptide from binding in multiple binding modes (registers) by locking the peptide in the desired orientation. This is supported by the crystal structures of the complexes and experiments in solution with single-molecule Förster resonance energy transfer (FRET) which we report here. We also provide the structural basis for how dArmRPs can differentiate between very similar amino acids, namely isoleucine and leucine, which supports our design goal of creating sequence-specific peptide binders.

Furthermore, we show that this lock is "movable" to another position in the dArmRPs, from the end to the middle, further underlining the modularity concept. The peptide can thus also be successfully locked by having the redesigned pocket in an internal position. Two crystal structures illustrate that both the grafted and the optimized binding pockets can indeed be freely moved to an internal position while still recognizing their target sequence.

With our design strategy, we provide a basis to lock the peptide into a specific binding register on the otherwise very repetitive nature of dArmRPs. This is not only a prerequisite for selecting sequence-specific recognition sites by common in vitro selection techniques, but it also opens

up new possibilities for the further computational design of new specific binding pockets for dArmRPs.

RESULTS AND DISCUSSION

A binding lock to prevent multiple binding modes. Currently, modular binding of designed ArmRPs has been demonstrated for positively charged peptides or alternating arginines and lysines[15]. Recently published crystal structures of regularized, consensus dArmRPs with bound peptides confirmed the modular binding mode, where each Arg-Lys dipeptide unit is recognized by one internal repeat[15]. However, the repetitive nature of both the peptide and the dArmRP enables several binding modes, i.e., register shifts and flips of the peptide in the binding pocket[19]. Multiple binding modes would hamper the development of new specificities by selection, as the randomized pocket on the dArmRP could be occupied by different amino acids in the different registers (Henning *et al.*, unpublished data).

One approach to address this problem is by breaking the symmetry and repetitiveness of the peptide and the binding pocket by creating a pocket with an orthogonal binding preference. Here, we achieved this goal by grafting a hydrophobic binding pocket observed in β catenin on a dArmRP, thus locking the peptide carrying the complementary sequence in a specific binding mode. The armadillo repeat domain of β catenin binds a variety of ligands, among them intrinsically disordered domains of transcription factors, which is a consequence of its diverse roles in the cell[29]. The central binding pocket P8 of β-catenin is of special interest, as most of the bound ligands have a phenylalanine at this position, interacting very specifically with the pocket[17].

A sequence alignment of 12 β-catenin structures (including one structure of the homologous protein plakoglobin), complexed with 6 different ligands, revealed that this phenylalanine is always part of a hydrophobic motif, interacting with pocket P8 (SI Figure 1). We hypothesized that this pocket introduces a hydrophobic peptide-binding motif, making it sufficiently different from the polar arginine and lysine pockets, while still following the conserved binding mode. The binding motif for Leu-Ser-Phe (LSF) was chosen, based on an alignment of the most common bound ligands. Although isoleucine (52%) is more common than leucine (39.5%) at position P6, leucine was chosen, because the highest-resolution structure available bound the motif LVF (PDB-ID: 1I7W). Serine and valine are equally common (both 26%) at

position P7, and Ser was introduced to lower the hydrophobicity of the whole motif. Finally, phenylalanine was chosen as the most common amino acid at position P8.

We inspected the residues in β-catenin that interact with this peptide motif and grafted the respective amino acids onto the binding surface of the dArmRPs to generate a binding pocket with a new specificity (Figure 2 A). The pocket was moved to the N-terminus of the dArmRP, to make it compatible with selections in which the randomized pocket is located in the central repeats of the dArmRP. Hereafter, the grafted pocket will be called Lock 1, comprising the following amino acid changes: R33W, S36R, Q37N, Q71S, E72S, W75G and E114L.

To test the validity of our approach, we conducted a fluorescence anisotropy binding assay with the peptide sequence KRKRKAKLSF, which binds the dArmRP containing the Lock 1 pocket, and we determined a K_D of 83 ± 10 nM (Table 1). One alanine residue was introduced into the peptide sequence to reduce its overall binding affinity, as K_Ds below 1 nM are more difficult to determine accurately with this assay, and we wished to determine all affinities with the same system. Interestingly, the observed affinity is in the same range as previously reported for $(KR)_4$ peptides binding to a dArmRP of the same length, which suggests an interaction of the KLSF peptide motif with the new pocket with a strength of approximately one KR unit[15].

To analyze the interactions in more detail, we co-crystallized a $(KR)_4$KLSF peptide with the Lock 1 dArmRP containing five internal repeats and the N-terminal grafted pocket. Indeed, the crystal structure shows a specific interaction between the LSF peptide motif and the Lock 1 pocket (Figure 2 B and C). The asymmetric unit consists of two dArmRPs, both binding the peptide.

One complex shows a good agreement with the model of the grafted pocket (Figure 2 A). The peptide lodges slightly deeper in the pocket than in the model, with both leucine and phenylalanine binding in the new pocket, stabilized by a cation-pi stacking between Arg36 of the dArmRP and phenylalanine. The exact same interaction between Arg36 and the phenylalanine was also aimed for in the grafted model, with a further stabilization of the hydrophobic patch by the leucine. Additionally, Arg36 also forms a salt bridge to the C-terminus of the peptide. It might be noted that between the second dArmRP and the corresponding LSF motif, a crystal contact was formed, thereby pulling the LSF motif out of its pocket.

The success of the grafting can be explained by the hydrophobic nature of the pocket. Only the hydrophobic interactions were reproduced in the crystal structure, but not the ionic interactions:

the newly formed interaction to the C-terminus was not predicted, but is a result of the peptide lodging deeper in the whole pocket, thereby allowing the new interaction to be formed. Our design protocol only rewards total interaction energy, and we have not attempted at this point to break it down into individual components during design. The overall success of this design is also in line with the analysis of computational interface designs by Stranges and Kuhlman[21], who suggested that hydrophobic designs are still more likely to succeed than highly polar ones.

Computational design increases the affinity of the grafted pocket. In the next step, we aimed to improve the specificity and binding affinity of the Lock 1 pocket. An alanine scan of the peptide revealed that the leucine residue contributed most to the binding, but the serine and phenylalanine did so only marginally (Table 1). To improve the affinity, we used the crystal structure of the grafted pocket as a starting point for a computational redesign using the RosettaScripts interface design protocol[30,31].

Based on the idea that the whole surface of the Lock 1 pocket can potentially contribute to binding, we allowed mutations in a sphere around 12 Å. Additionally, positions W33 and N37 were changed back to the original amino acids (R33 and Q37) of the dArmRP, as they did not interact significantly with the peptide in the crystal structure. Different short peptide motifs were considered as targets for the design: LSF, KLSF and AKLSF were used as binding partners for the dArmRP. The sequence AKLSF was finally chosen, in order to decrease side-effects of the neighboring arginine on the outcome of the design, which would have led to the design of an RKLSF pocket. The output design models were analyzed with the interface analyzer in Rosetta and the model with the lowest computed Rosetta delta-delta G (ddG) energy value from each setup was chosen for expression and affinity measurements[30]. All models showed only a few mutations, ranging from six to eight, of which five were conserved in all designs (SI Figure 2). After measuring the affinities to the peptide KRKRKAKLSF, we found that all designs bind better to the peptide than the original pocket (Table 2). The pocket from the redesign with the AKLSF peptide has the best affinity (18 ± 2 nM), always measured in the context of KRKRKAKLSF, a four-fold increase when compared to the original pocket (Table 2).

The crystal structure of the new design (Lock 2, Figure 3 A, B) with the bound peptide KRKRKAKLSF demonstrated that the peptide is bound in a similar way as before. The asymmetric unit contains six molecules, all of which bind the peptide in exactly the same way. Among them, chain B shows the best electron density for the LSF moiety. There, both leucine and serine are well defined, but the terminal phenylalanine has an elevated B-factor, indicating

flexibility of the sidechain. The mutations R36Q and N37Q make additional hydrogen bonds to the backbone and to the serine side chain, respectively, whereas S40M increases the contact area between the peptide and the dArmRP. The mutations W117I and L114F increased the hydrophobic interaction of the peptide, also the threonine is involved in a hydrogen bond to Q37.

L114F increase the hydrophobic interaction of the peptide and provide a better hydrophobic pocket for the leucine. Furthermore, the interaction of the peptide C-terminus to R36 is shifted to R33 (Figure 3 A, B). The mutation S78A is not directly involved in peptide binding, but has been observed in all Rosetta interface designs of dArmRPs so far and might lead to a better packing of the helices.

Further redesign of the peptide additionally increases the affinity. The structure of the dArmRP with new Lock 2 pocket in complex with the KRKRKAKLSF peptide showed that this alternative solution to the original pocket can bind the peptide with higher affinity. Yet, key interactions like the cation-pi stacking between Arg36 and the phenylalanine of the peptide were not observed in the new design. For this reason, and because of the flexibility of the terminal phenylalanine, which might not fit perfectly in the pocket, we decided to do a second round of computational design on Lock 2, this time changing the peptide sequence instead of the pocket, to investigate whether another peptide sequence would bind even better to the new lock.

The design was carried out analogously to the first round and a predominant outcome of the design was the sequence LTW, with 82 out of 100 designed sequences. Besides LTW, especially LSW (10%) and ITW (6%) were predicted to bind. We decided to experimentally test the top five sequences for their ability to bind to the new pocket.

Interestingly, most of the new sequences showed a higher affinity (lower dissociation constant K_D) than the original LSF motif, and one sequence showed the highest affinity among all other peptide motifs. The KRKRKAKITW peptide binds with an affinity of 7 ± 0.3 nM, a more than 2-fold decrease of the K_D (Table1). The crystal structure, which was obtained in the same condition and space group as for the LSF peptide, explains this by showing that ITW fits deeper into the pocket than LSF (Figure 3 C, D). The electron density is again not identical among the six protein-peptide complexes of the asymmetric unit, but all of them exhibit the same binding. Chain D was used for further analysis. The isoleucine is bound in a similar way as the leucine

before. Furthermore, a cation-pi interaction between the tryptophan and R33 is observed. Like the serine of the peptide, also the threonine is involved in a hydrogen bond to Q37.

These interactions, combined with the fact that with isoleucine and tryptophan two larger hydrophobic residues were introduced, can explain the higher affinity of the ITW peptide compared to the LSF peptide. This observation is also consistent with the increase in shape complementarity (SC) between peptide and dArmRP. It should be noted that for small ligands like peptides, the SC can be affected by edge effects, but when comparing the three similar interfaces, this effect should be minor[32]. The original pocket (Lock 1) binding the LSF motif has a SC of 0.76, which is comparable to typical antibody-peptide complexes[32]. Importantly, both LSF (0.809, chain B) and ITW (0.848, chain D) peptides show a higher SC when bound to the new pocket (Lock 2).

The Lock 2 pocket is specific for isoleucine over leucine. By testing all the different peptide motifs, we found a remarkable difference in affinity between the leucine and isoleucine variants. Whereas the LTW peptide binds with an affinity of 23 ± 1 nM, ITW binds with 7 ± 0.3 nM, which is a difference of a factor three between these very similar amino acids in a pocket that was not specifically designed for isoleucine. By comparing the structures with both the LSF and the ITW peptides, it becomes obvious that isoleucine fits better into the optimized pocket (SI Figure 3). Leucine has a branch at the Cγ-atom that leaves a hole opposite Gly75 which is filled by isoleucine, which has a branch at the Cβ-atom. Furthermore, the branch at the Cγ-atom of leucine could also result in a clash with I117. Consequently, the observed rotamer of this leucine residue is the only one that is not involved in any potential clashes between the sidechain and the surrounding amino acids. This observation is very encouraging for the design of sequence-specific pockets.

Both design rounds resulted in a decreased off-rate. To understand the nature of the interaction of peptides with dArmRPs better, we developed a stopped flow Förster resonance energy (FRET) assay to determine the binding and dissociation rates for both the dArmRPs (containing Lock 1 and Lock 2) and the peptides (KRKAKRKLSF and KRKAKRKITW). For this purpose, the dArmRP and the peptide were fused to the fluorescent proteins mCherry and sfGFP, respectively. The association was probed by following the increase in mCherry fluorescence upon binding, when sfGFP was excited. The dissociation rate constant was determined in a competition assay, where a pre-formed dArmRP-mCherry/peptide-sfGFP complex was rapidly mixed with an excess of unlabeled peptide and the decrease of acceptor fluorescence was monitored.

Table 3 and SI Figures 4-7 show that all dArmRP-peptide interactions share the same very high association rate constant (>10^8 M^{-1} s^{-1}), consistent with an electrostatically driven interaction of an unstructured peptide without rotational restrictions. For the interaction of $Y_{III}M_5A_{II}$ with $(KR)_5$ (k_{on} = (9.7 ± 1.3)·10^8 M^{-1} s^{-1}), a dArmRP with five identical binding repeats each recognizing a lysine-arginine (KR) dipeptide unit, an even higher association rate than the dArmRPs with the new pockets (by a factor of three) is seen.

These rates are close to the limit set by diffusion[33], and probably accelerated by electrostatic attraction between the positively charged peptide and the negatively charged dArmRP, which explains also the higher k_{on} of the $Y_{III}M_5A_{II}$:$(KR)_5$ complex compared to the LSF- or ITW-containing constructs. It should be noted that the fluorescent fusion proteins carry only a miniscule net charge of -3 at the pH of the experiment. Furthermore, the constant KRKRKAK-motif of all Lock-peptides may contribute the majority to the fast association, which explains why all rate coefficients are nearly identical. The dissociation rate constants of dArmRPs containing Lock 1 and Lock 2 show larger differences, ranging from fast dissociation of the Lock 1:KRKRKAKLSF peptide complex (46.6 ± 3.1 s^{-1}) to 30 fold slower dissociation in the optimized Lock 2:KRKRKAKITW peptide complex (1.6 ± 0.0 s^{-1}). This observation strongly indicates that the increase in binding affinity is mainly driven by a decrease of k_{off}, which we attribute to the additional interactions within the peptide lock.

The designed pocket prevents sliding in solution. The different binding registers between peptide and dArmRP, whose elimination was the driving force behind this study, were so far only observed in crystal structures. It might thus be possible that these register shifts are induced by crystal packing forces. Therefore, we conducted single-molecule FRET experiments to probe if register shifts also occur in solution[34]. Two peptide-dArmRP complexes were investigated: First, the symmetric $Y_{III}M_5A_{II}$:$(KR)_5$ complex, where register shifts have been observed in the crystal structure; second, the Lock 2:KRKRKAKITW complex, where the peptide should be locked into one preferential binding register. In both cases, the peptide was labeled with a donor dye (Cy3B) at an N-terminal cysteine residue, and with an acceptor dye (CF660R) at position 51 (mutated to cysteine) of the dArmRP. The measured transfer efficiency is directly related to the distance between the two dyes, which will be affected by register changes (Figure 4). Single-molecule FRET captures "distance snapshots" (time resolution of ≈1 ms) of individual molecules as they diffuse through the laser focus, so each binding register will generate a distinct transfer efficiency peak if it is populated in solution.

The transfer efficiency (E) histograms of the $Y_{III}M_5A_{II}$:$(KR)_5$ and the Lock 2:KRKRKAKITW complex are shown in Figure 4. Both histograms share a very similar average peak position. However, the peak is much broader for $Y_{III}M_5A_{II}$:$(KR)_5$, indicating that in this complex, a broader distribution of states is populated. For a quantitative description, we first conducted a recurrence analysis of single particles (RASP[35]) to test how many sub-species are required to describe the transfer efficiency histograms (see SI Figure 8). RASP revealed that at least three species are contributing to the histogram of $Y_{III}M_5A_{II}$:$(KR)_5$, potentially corresponding to one preferred binding mode and sliding of the peptide to either side of the dArmRP.

We hypothesized that the optimal binding mode is populated to a larger extent in the Lock 2:KRKRKAKITW complex, and the two shifted binding modes contribute less, leading to a narrower E distribution compared to $Y_{III}M_5A_{II}$:$(KR)_5$. Nonetheless, for each binding register, the underlying inter-dye distance distributions should be the same in both complexes, as the peptide backbone is bound in the same way. Therefore, we fitted both histograms with three populations using photon distribution analysis (PDA[36]), which calculates the shape of the transfer efficiency peaks by taking into account shot noise, the statistical fluctuation of recorded numbers of donor and acceptor photons about their mean values. We obtained a very good fit to the recorded E histograms by assuming the underlying E distributions to be very narrow and sharing the same peak positions for both dArmRP-peptide complexes (see Figure 4). The resulting fit suggests that the optimal binding mode (shaded in blue) is occupied only to 50% in the $Y_{III}M_5A_{II}$:$(KR)_5$ complex. In the Lock 2:KRKRKAKITW complex, however, it dominates the distribution with 68% occupancy. We note that this value increases to 85% when we allow slight peak broadening beyond shot noise (a common observation in many single-molecule FRET studies[37]; see SI Figure 9), so the value of 68% likely represents a lower limit for the population of the optimal state in the Lock 2:KRKRKAKITW complex.

To compare the FRET results with the atomic models for the dArmRP-peptide complexes, we calculated the transfer efficiencies expected from the complex structures. The $Y_{III}M_5A_{II}$:$(KR)_5$ structure (PDB-ID: 5AEI) served as model for the optimal binding mode, and the sliding states (register shifts) were modeled with Coot[38]. The dyes can sample a large range of inter-dye distances as they are attached to the protein by long flexible linkers. For each structure we calculated the geometrically accessible volumes for both dyes[39] and estimated the mean transfer efficiencies, assuming fast inter-dye distance fluctuations. The analysis yields transfer efficiencies of 0.74, 0.83 and 0.87, for the -1, optimal, and +1 states, close to the measured values (0.76, 0.84 and 0.92; typical accuracy ±0.03). In conclusion, the single-molecule FRET

experiments support the presence of sliding in solution for the symmetric $Y_{III}M_5A_{II}$:$(KR)_5$ complex, and suggest that it is significantly reduced for the restrained Lock 2:KRKRKAKITW complex, in line with the crystallographic results.

The binding pockets can be transferred to an internal position. The peptide lock, located in the N-terminal region of the ArmRP, was designed to bind the C-terminal amino acids of the peptide. Such a terminal peptide lock would prohibit peptide register shifts, and can be used with a randomization strategy in the middle of the protein in selections for a new amino acid side chain located in the middle of the peptide. To test the modularity of our system we moved the module, recognizing the new amino acids (K)LSW or (K)ITW, orthogonal to KRKR.

To test whether the pocket can be moved to the middle of the dArmRP, we constructed two different internal locks. Based on the structural similarity of the N-cap and the internal repeats, we took the structure PDB-ID: 5MFM, a dArmRP with six internal repeats binding to a $(KR)_5$ peptide and grafted the binding residues of the new pockets onto the internal repeats two to four, keeping to their relative positions like at the N-terminus. For the Lock 1 pocket, which binds to LSF, we knew that only the leucine and the phenylalanine were involved in binding. Therefore, we designed the new peptide with a lysine at the position of the serine, having the sequence KRKRKLKFKR. For the improved pocket, also the threonine displayed an interaction in the crystal structure. Therefore, we tested the sequence KRKAKITWKR, analogously to the terminal AKITW motif from the original design.

Both new dArmRPs bind to the respective peptides with an affinity of 298 ± 45 nM (internal Lock 1) or 102 ± 19 nM (internal Lock 2) and an alanine scan on the lock binding residues of the peptide revealed that the lock is indeed involved in binding (Table 4). However, the overall affinity is lower than for the initial lock designs (83 ± 10 nM for Lock 1 and 7 ± 0.3 for Lock 2), where the respective amino acids and the pockets were placed at the termini of the peptide and the dArmRP.

The structures of the internal pockets in complex with their target peptides show that the peptides are bound differently on the surface of the dArmRP, because of their different designs (Figure 5). The internal Lock 1 binds the peptide in the conserved binding mode, with every second peptide bond being involved in bidentate hydrogen bonds to the conserved N37 ladder. Additionally, also all arginines and lysines are bound in their respective pockets and the side-chains from the new peptide fit into the internal pockets. This binding mode is therefore as expected, since the peptide backbone was kept in place by the interaction to the N37 ladder.

Yet, it deviates slightly from the binding mode of the terminal Lock 1 pocket. There, the C-terminus of the peptide is free to move deeper into the hydrophobic pockets, which is not possible in the internal position, as the backbone is fixed by binding to N37.

In contrast to the results with Lock 1, the internal Lock 2 exhibits a different binding between a terminal and an internal location. The peptide is only bound up to the ITW moiety. The ITW moiety shows a similar conformation in the internal position when compared to ITW in the terminal position, although the terminal tryptophan is not bound as deep into the pocket as in the terminal position. However, no clear electron density is visible for the additional arginine and lysine residues which succeed the ITW sequence, indicating that they do not bind into their respective pockets, and the whole peptide stretch is effectively locked into position by just the ITW moiety. Furthermore, the interaction of the backbone to the conserved N37 is not possible in the internal Lock 2 pocket. At this position, the N37 was mutated to a glutamine, which interacts with the threonine of the ITW motif and not with the backbone of the peptide. By this interaction, the whole peptide is locked into this binding register at the expense of constraining the peptide backbone to the protein. This also explains the lower affinity, compared to a peptide of the same composition and length but with a terminal ITW motif, where the lysine and arginine can still bind.

Conclusions

Modular binding scaffolds have the potential to overcome time-intense and expensive selections for individual peptide binders, as a preselected set of specific binding modules can be reassembled to bind arbitrary peptide targets[17,18]. We had observed that register shifts, i.e., multiple binding modes, can exist, which could interfere with the desired specific recognition of individual amino acids[19]. Furthermore, a specific binding register is mandatory during selections, where a new target amino acid is facing a randomized pocket. Here we show that this problem of peptide register shifts of repeating units can be overcome by introducing a designed peptide lock. We created the new binding site neither by complete de novo design nor by selection (as the latter would be dependent on already having a functional lock), but we grafted a natural hydrophobic binding pocket of β-catenin onto our dArmRP scaffold. In our study, we reached the following conclusions:

(i) We found that one of the two designed pockets (Lock 2) was able to differentiate between isoleucine and leucine by a factor of three (ITW:LTW). This result shows that dArmRPs have the potential to differentiate between very similar amino acids, and this finding is important for the further design for specificity.

(ii) By developing a stopped flow based FRET assay we noted very high k_{on}. This can be explained by the mechanism of binding, which involves an elongated and short unstructured peptide stretch, furthermore involving charge interactions between protein and peptide. This increases the likelihood of successful collisions, as observed in other systems[40,41].

(iii) Single-molecule FRET studies indicated multiple binding modes in solution for the repetitive system, and thus showed that different binding modes (registers) are not just crystallization artefacts but do also exist in solution. In contrast, a much larger fraction of the locked peptide is bound in the desired register, probably closer to 85%, if we consider that there may be a slight peak broadening beyond shot noise. This underlines that we have successfully created a peptide lock with only one predominant binding mode.

(iv) We found that the two new pockets can be moved from an N-terminal to an internal position, thereby creating the first designed modular peptide binders, in which the modules can be easily interchanged. Both the biochemical and structural analysis of these shuffled dArmRP showed an interaction to the respective peptide, with the structures displaying no peptide register shift, as desired, with the same interactions (Lock 2) as at the terminus.

In summary, we describe the design of a hydrophobic binding pocket for dArmRP repeat proteins to solve a particular problem, namely the sliding and flipping of repetitive peptides on repetitive dArmRPs. A combination of grafting and computational redesign resulted in a sequence-specific pocket that is orthogonal to those recognizing positively charged amino acids. The new design can discriminate between two very similar amino acids, isoleucine and leucine. With this study we also give a first example of a workflow for the computational design of pockets for dArmRPs. Furthermore, we show that this new pocket prevents multiple binding modes and can be freely moved to an internal position. This demonstrates the potential of dArmRPs as modular peptide binders. The collection of pockets for other amino acids will now be extended by a combination of selection and de novo design, testing new algorithms, score functions and different constraints.

METHODS

Cloning, Protein Expression and Purification. Cloning was carried out as described[9,13]. DNA fragments coding for the respective proteins were ordered from IDT or Genewiz and subcloned into a pQE30LIC_3C-based vector as described[15]. Protein expression and purification were done as described previously[16], or alternatively, expression was done in auto-induction medium for 15 hours at 25 °C[42].

Crystallization and structure determination. Proteins were concentrated to 20-150 mg/mL in 10 mM Tris/HCl pH 8, 100 mM NaCl using Amicon centrifugal concentrators (Amicon Ultra Centrifugal Filters, Merck Millipore) and a 1.5-fold molar excess of peptide was added to the Lock 1 protein. Screening for suitable crystallization conditions was done in commercially available 96-well sparse matrix screens (Hampton Research and Molecular Dimensions). Per condition, the mother liquor was mixed in three different ratios (1:1, 2:1, 3:1 or 5:1) with the protein in 300-400 nl drops and 75 µL of reservoir solution, plates were incubated at 4 °C. For structure 6S9L a fine-screen was set up based on the initially found conditions by changing the pH along the column and the precipitant along the row in a 96 well format.

Prior to data collection, crystals were flash-frozen in liquid nitrogen directly or after incubation for 10 seconds in mother liquor containing 20% ethylene glycol. Diffraction data were collected at beamline X06SA (Paul Scherrer Institute, Villigen, Switzerland) equipped with an Eiger 16M detector (Dectris) at 1 Å. Crystallization conditions and data collection and refinement statistics are summarized in SI Table 1. Data were processed with XDS, XSCALE and XDSCONV[43]. Phases were determined by molecular replacement using PHASER[44] with structures of dArmRPs containing five or six internal repeats. Model building was done in Coot[38] and refinement using REFMAC5[45], PHENIX refine[46] and BUSTER[47]. The final resolution of the datasets was determined using paired refinement in pdb_redo[48] based on[49].

Determination of binding constants. All experiments were conducted in PBS, supplied with 0.01% Tween 20. Fluorescence anisotropy data were measured according to[15] with 2-3 measurements per point. Binding kinetics were determined using a FRET-based binding assay. Increasing concentrations of mCherry-dArmRP fusions were mixed with a constant concentration of sfGFP-peptide fusions in a 1:1 volume ratio in 120 µL in a PiStar-180 stopped-flow fluorimeter (Applied Photophysics Ltd) equipped with a mercury-xenon lamp. sfGFP was excited using a wavelength of 436 nm and a bandwidth of 10 nm. Fluorescence was collected

through a 590 nm long-pass filter. For obtaining k_{on}, three binding curves at different mCherry-dArmRP concentrations were recorded, a single-exponential rise was fit to the data and the resulting rates were plotted against the concentration of mCherry-dArmRP (120 nM, 160 nM, 200 nM). The final k_{on} was then determined by fitting a linear function to these data points. An additional point at 80 nM was also recorded but not used for the fit, as the concentration was only two times above the peptide-sfGFP concentration. For k_{off}, the peptide-protein complex was pre-incubated and mixed with two different concentrations of competitor, non-labelled dArmRP. A single-exponential decay was fit to the data.

Single molecule FRET experiments. For labelling of dArmRPs, $Y_{III}M_5A_{II}$ (a dArmRP with 5 internal KR binding repeats; Y_{III} and A_{II} are the third- and second-generation capping repeats, respectively) and Lock 2, a cysteine was introduced at position D51 via site-directed mutagenesis. Protein expression and purification was done as described above with 1 mM 2 mercaptoethanol (Pierce) in all purification buffers. Proteins were further purified via size exclusion chromatography in PBS, 1 mM DTT and a consecutive anion exchange chromatography on a MonoQ 5/50 GL column to remove the DTT in 20 mM NaH_2PO_4/Na_2HPO_4, pH 7.2. Proteins were eluted with a gradient from 0-1000 mM NaCl. Following the elution, the proteins were labelled using CF660R-maleimide dye (Biotium) in 1.3-fold molar excess for 16 h at 4 °C. After labelling, the excess dye was removed with a second anion exchange step. Peptides (CSAGGKRKRKRKR and CSAGGKRKRKAKITW) were ordered from LifeTein as lyophilized powder and dissolved in 20 mM NaH2PO4/Na2HPO4, pH 7.2, 1 mM TCEP. They were further purified by reverse phase-HPLC on a Reprosil Gold 200 C18 column (Dr. A. Maisch GmbH,), using a H_2O + 0.1% (v/v) TFA/acetonitrile gradient. Fractions containing the full-length peptide were pooled and lyophilized. For labelling, the peptides were dissolved in 100 mM KH_2PO_4/K_2HPO_4, pH 7.0 and a 1.3-fold excess of Cy3B-maleimide (GE Healthcare) was used for labelling the N-terminal cysteine. The reaction was incubated at room temperature for 3 h, followed by a second HPLC step to remove excess dye. The labelled peptides were again lyophilized and stored at 20 °C until further use. For the measurement, they were redissolved in 50 mM NaH_2PO_4/Na_2HPO_4, pH 7.0, supplied with 0.001% (v/v) Tween 20.

Measurements were recorded for 1 h with Cy3B-labeled peptide concentrations of 50 pM and CF660R-labeled dArmRP concentrations of 5 nM, in 50 mM NaH_2PO_4/Na_2HPO_4, pH 7.0, supplied with 0.001% (v/v) Tween 20. A plastic sample chamber (μ-slide Angiogenesis, ibidi) was used to minimize surface adhesion.

All single-molecule experiments were conducted on a custom-built confocal instrument equipped with a 532 nm continuous-wave laser (LaserBoxx LBX-532-50-COL-PP, Oxxius) to excite the donor dye[50]. Fluorescence photons were collected through a high numerical aperture objective (UPlanApo 60×/1.20-W, Olympus), and subsequently separated from the scattered photons with a triple-band mirror (zt405/530/630rpc, Chroma) and split onto four channels according to their wavelength and polarization. Dichroic mirrors were used to separate donor and acceptor emission (T635LPXR, Chroma). Donor photons were filtered with ET585/65m bandpass filters (Chroma) before detection on one of two τ-SPAD avalanche photodiodes (PicoQuant). Acceptor photons were filtered with LP647RU long pass filters (Chroma) and detected with SPCM-AQRH-14 single-photon avalanche diodes (Perkin Elmer). The arrival time of every photon was recorded; the photon time trace was binned at 0.75 ms and all bins with at least 80 photons in them were retained as a burst. Identified bursts were corrected for the quantum yield of the dyes, the detection efficiency of the detectors, crosstalk and direct excitation of the acceptor[51]. Bursts during which acceptor photobleaching likely occurred were eliminated[52]. The bursts were binned in a transfer efficiency (E) histogram according to the corrected number of photons detected in the donor (n_D) and acceptor (n_A) channels: $E = \frac{n_A}{n_A + n_D}$

RASP analysis of the transfer efficiency histogram was conducted as described before[35]. Burst pairs separated by less than 6 ms were selected for the analysis. From the time correlation of the bursts we calculate that after 5 ms the probability that the bursts originate from the same molecule is still 85%. We filter for groups of burst pairs where the transfer efficiency of the first burst is in an interval of ΔE = E2-E1 = 0.05, ranging from E1 of 0.65 to 1 in steps of 0.025. The second bursts of each pair are used to construct "pure" species histograms. Those 15 resulting histograms were normalized to an area of 1, and globally fitted with one to four Gaussian peak functions to calculate the χ^2 value used to assess the quality of the fit.

Probability distribution analysis (PDA[36]) was used to analyze transfer efficiency histograms. PDA assumes an intrinsic distribution of experimental transfer efficiencies, which is further broadened by shot noise due to the limited number of photons detected per burst. Shot-noise broadening is modeled using the photon statistics of the experimental burst size distribution. Here we applied PDA, assuming that the peptide-dArmRP-complex can exist in three states that have the same intrinsic transfer efficiency distributions for both the $Y_{III}M_5A_{II}$:$(KR)_5$ complex as well as the Lock 2:ITW complex. Unlike the original PDA approach[36], however, we calculate the intrinsic distribution not from a hypothetical normal distribution of inter-dye

distances. Instead, we model the intrinsic transfer efficiency distribution, $p(E)$, for each subpopulation directly in transfer efficiency space in the form of a beta distribution,

$$p(E) = \frac{(1-E)^{(1-\varepsilon)\nu-1} E^{\varepsilon\nu-1}}{\mathrm{B}(\varepsilon\nu,(1-\varepsilon)\nu)}$$

where $\mathrm{B}(x,y)$ is the beta function, $\varepsilon = \langle E \rangle$ is the mean of the distribution, and its variance is given by $\sigma^2 = \varepsilon(1-\varepsilon)/(1+\nu)$ with the parameter $\nu > 0$. In Figure 4, we assumed an infinitely narrow intrinsic transfer efficiency distribution ($\nu \to \infty$), in SI Figure 8 we left ν as a free fit parameter but constrained it to be identical for all three subpopulations.

Geometrically accessible volumes of dyes were calculated from different models of the $Y_{III}M_5A_{II}$:$(KR)_5$ complex (models were generated in Coot based on PDB-ID: 5AEI), which served as model structures for the ligand-bound and ligand-free states, respectively. All calculations were performed using the freely available "FRET Positioning and Screening" software (http://www.mpc.hhu.de/en/software/fps.html), in which Cy3B and CF660R were parameterized with the following parameters (linker length/width/dye radii in Å): 37/4.5/(3.4/8.2/3) for Cy3B, and 19/4.5/(8.1/4.2/2.1) for CF660R. The long linker length for Cy3B arises because we assume the five N-terminal amino acids of the peptide to be flexible, as they have no binding pocket on the dArmRP (assuming a stretched chain; adding 19 Å to the Cy3B linker length of 18 Å). To take into account the rapid fluctuations in inter-dye distance that occur on timescales similar to the excited state lifetime of the donor, we modeled the dynamics as diffusive motion in the potential of mean force corresponding to the inter-dye distance distribution from the accessible volume calculations using the diffusion coefficient of the free dyes[53].

In RASP the single molecule data are filtered for pairs of photon bursts that were measured so close in time that they most likely were emitted from one and the same molecule visiting the confocal volume twice. As long as the molecule does not change its conformation, both bursts will sample the same underlying transfer efficiency distribution of a subpopulation. This way, we are able to reconstruct individual subpopulations from a histogram with overlapping peaks (see above).

Accession Codes. Protein models have been deposited at the Protein Data Bank under the accession codes 6S9L (Lock 1:$(KR)_4$KLSF), 6S9M (Lock 2_$(GS)_6$_KRKRKAKLSF), 6S9N (Lock 2_$(GS)_6$_KRKRKAKITW), 6S9O (internalLock_2_$(GS)_6$KRKRKLKFKR) and 6S9P (internalLock_2_$(GS)_6$_KRKAKITWKR).

ASSOCIATED CONTENT

Supporting Information. SI Figure 1, ligands binding to β-catenin; SI Figure 2, alignment of designed proteins; SI Figure 3, surface view of peptides binding Lock 2; SI Figure 4-7, fit of kinetic binding constants; SI Figure 8, RASP analysis of $Y_{III}M_5A_{II}$:$(KR)_5$ complex; SI Figure 9 transfer efficiency histogram1, of $Y_{III}M_5A_{II}$:$(KR)_5$ complex and Lock 2:ITW complex; SI Table 1, data collection and refinement statistics of shown protein structures; Rosetta scripts code for redesigns; This material is available free of charge via the Internet at http://pubs.acs.org."

AUTHOR INFORMATION

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by the Schweizerische Nationalfonds BRIDGE Program grant 20B2-1_176535 and the EU FET-OPEN program grant Pre-ART (to AP).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank Céline Stutz-Ducommun and Beat Blattmann from the UZH Protein Crystallization Center for setting-up crystallization experiments and the staff of beamlines X06DA an X06SA at the Swiss Light Source (Paul Scherrer Institut, Würenlingen, Switzerland) for technical support.

ABBREVIATIONS

dArmRP, designed Armadillo repeat protein, nArmRP, natural Armadillo repeat protein, DARPin, designed Ankyrin repeat protein, GFP, green fluorescent protein

REFERENCES

(1) Petsalaki, E., and Russell, R. B. (2008) Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr. Opin. Biotechnol. 19*, 344–350.

(2) Xue, B., Dunker, A. K., and Uversky, V. N. (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn. 30*, 137–149.

(3) Zahnd, C., Wyler, E., Schwenk, J. M., Steiner, D., Lawrence, M. C., McKern, N. M., Pecorari, F., Ward, C. W., Joos, T. O., and Plückthun, A. (2007) A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *J. Mol. Biol. 369*, 1015–1028.

(4) Conti, E., and Kuriyan, J. (2000) Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin ?? *Structure 8*, 329–338.

(5) Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. (1998) Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell 94*, 193–204.

(6) Huber, A. H., and Weis, W. I. (2001) The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell 105*, 391–402.

(7) Parmeggiani, F., Pellarin, R., Larsen, A. P., Varadamsetty, G., Stumpp, M. T., Zerbe, O., Caflisch, A., and Plückthun, A. (2008) Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol. 376*, 1282–1304.

(8) Varadamsetty, G., Tremmel, D., Hansen, S., Parmeggiani, F., and Plückthun, A. (2012) Designed Armadillo Repeat Proteins: library generation, characterization and selection of peptide binders with high specificity. *J. Mol. Biol. 424*, 68–87.

(9) Madhurantakam, C., Varadamsetty, G., Grütter, M. G., Plückthun, A., and Mittl, P. R. E. (2012) Structure-based optimization of designed armadillo-repeat proteins. *Protein Sci. 21*, 1015–28.

(10) Alfarano, P., Varadamsetty, G., Ewald, C., Parmeggiani, F., Pellarin, R., Zerbe, O., Plückthun, A., and Caflisch, A. (2012) Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy. *Protein Sci. 21*, 1298–1314.

(11) Watson, R. P., Christen, M. T., Ewald, C., Bumbak, F., Reichen, C., Mihajlovic, M.,

Schmidt, E., Güntert, P., Caflisch, A., Plückthun, A., and Zerbe, O. (2014) Spontaneous self-assembly of engineered armadillo repeat protein fragments into a folded structure. *Structure 22*, 985–995.

(12) Ewald, C., Christen, M. T., Watson, R. P., Mihajlovic, M., Zhou, T., Honegger, A., Plückthun, A., Caflisch, A., and Zerbe, O. (2015) A combined NMR and computational approach to investigate peptide binding to a designed armadillo repeat protein. *J. Mol. Biol. 427*, 1916–33.

(13) Reichen, C., Hansen, S., Forzani, C., Honegger, A., Fleishman, S. J., Zhou, T., Parmeggiani, F., Ernst, P., Madhurantakam, C., Ewald, C., Mittl, P. R. E., Zerbe, O., Baker, D., Caflisch, A., and Plückthun, A. (2016) Computationally Designed Armadillo Repeat Proteins for Modular Peptide Recognition. *J. Mol. Biol. 428*, 4467–4489.

(14) Reichen, C., Madhurantakam, C., Plückthun, A., and Mittl, P. R. E. (2014) Crystal structures of designed armadillo repeat proteins: Implications of construct design and crystallization conditions on overall structure. *Protein Sci. 23*, 1572–1583.

(15) Hansen, S., Tremmel, D., Madhurantakam, C., Reichen, C., Mittl, P. R. E., and Plückthun, A. (2016) Structure and energetic contributions of a designed modular peptide-binding protein with picomolar affinity. *J. Am. Chem. Soc. 138*, 3526–3532.

(16) Hansen, S., Ernst, P., König, S. L. B., Reichen, C., Ewald, C., Nettels, D., Mittl, P. R. E., Schuler, B., and Plückthun, A. (2017) Curvature of designed armadillo repeat proteins allows modular peptide binding. *J. Struct. Biol. 201*, 108–117.

(17) Reichen, C., Hansen, S., and Plückthun, A. (2014) Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins. *J. Struct. Biol. 185*, 147–62.

(18) Ernst, P., and Plückthun, A. (2017) Advances in the design and engineering of peptide-binding repeat proteins. *Biol. Chem. 398*, 23–29.

(19) Hansen, S., Kiefer, J. D., Madhurantakam, C., Mittl, P. R. E., and Plückthun, A. (2017) Structures of designed armadillo repeat proteins binding to peptides fused to globular domains. *Protein Sci. 26*, 1942–1952.

(20) Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. a, Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y.-E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo,

M., Mentzer, S., Popović, Z., Havranek, J. J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol. 487*, 545–74.

(21) Benjamin Stranges, P., and Kuhlman, B. (2013) A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci. 22*, 74–82.

(22) Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E.-M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science 332*, 816–21.

(23) Baran, D., Pszolla, M. G., Lapidoth, G. D., Norn, C., Dym, O., Unger, T., Albeck, S., Tyka, M. D., and Fleishman, S. J. (2017) Principles for computational design of binding antibodies. *Proc. Natl. Acad. Sci. 114*, 10900–10905.

(24) Mou, Y., Huang, P.-S., Hsu, F.-C., Huang, S.-J., and Mayo, S. L. (2015) Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl. Acad. Sci. 112*, 10714–10719.

(25) Sammond, D. W., Bosch, D. E., Butterfoss, G. L., Purbeck, C., Machius, M., Siderovski, D. P., and Kuhlman, B. (2011) Computational Design of the Sequence and Structure of a Protein-Binding Peptide. *J. Am. Chem. Soc. 133*, 4190–4192.

(26) Lewis, S. M., and Kuhlman, B. A. (2011) Anchored Design of Protein-Protein Interfaces. *PLoS One 6*, 1–14.

(27) Liu, X., Taylor, R. D., Griffin, L., Coker, S.-F., Adams, R., Ceska, T., Shi, J., Lawson, A. D. G., and Baker, T. (2017) Computational design of an epitope-specific Keap1 binding antibody using hotspot residues grafting and CDR loop swapping. *Sci. Rep. 7*, 41306.

(28) Procko, E., Hedman, R., Hamilton, K., Seetharaman, J., Fleishman, S. J., Su, M., Aramini, J., Kornhaber, G., Hunt, J. F., Tong, L., Montelione, G. T., and Baker, D. (2013) Computational Design of a Protein-Based Enzyme Inhibitor. *J. Mol. Biol. 425*, 3563–3575.

(29) Valenta, T., Hausmann, G., and Basler, K. (2012) The many faces and functions of β-catenin. *EMBO J. 31*, 2714–2736.

(30) Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., and Baker, D. (2011)

RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS One* (Uversky, V. N., Ed.) *6*, e20161.

(31) Bender, B. J., Cisneros, A., Duran, A. M., Finn, J. A., Fu, D., Lokits, A. D., Mueller, B. K., Sangha, A. K., Sauer, M. F., Sevy, A. M., Sliwoski, G., Sheehan, J. H., DiMaio, F., Meiler, J., and Moretti, R. (2016) Protocols for Molecular Modeling with Rosetta3 and RosettaScripts. *Biochemistry 55*, 4748–63.

(32) Kuroda, D., and Gray, J. J. (2016) Shape complementarity and hydrogen bond preferences in protein–protein interfaces: implications for antibody modeling and protein–protein docking. *Bioinformatics 32*, 2451–2456.

(33) Smoluchowski MV. (1918) Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Zeitschrift für Phys. Chemie*.

(34) Selvin, P. R., and Ha, T. (2008) Single-Molecule Techniques: A Laboratory Manual. *Cold Spring Harb. Lab. Press. New York*.

(35) Hoffmann, A., Nettels, D., Clark, J., Borgia, A., Radford, S. E., Clarke, J., and Schuler, B. (2011) Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: recurrence analysis of single particles (RASP). *Phys. Chem. Chem. Phys. 13*, 1857–71.

(36) Antonik, M., Felekyan, S., Gaiduk, A., and Seidel, C. A. M. (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B 110*, 6970–8.

(37) Schuler, B., and Eaton, W. A. (2008) Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol. 18*, 16–26.

(38) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 486–501.

(39) Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., Goody, R. S., Gohlke, H., and Seidel, C. A. M. (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods 9*, 1218–1225.

(40) Schreiber, G., and Fersht, A. R. (1996) Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol. 3*, 427–431.

(41) Stone, S. R., Dennis, S., and Hofsteenge, J. (1989) Quantitative evaluation of the

contribution of ionic interactions to the formation of the thrombin-hirudin complex. *Biochemistry 28*, 6857–6863.

(42) Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif. 41*, 207–34.

(43) Kabsch, W. (2010) XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 125–132.

(44) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658–674.

(45) Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr. 67*, 355–367.

(46) Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr. 68*, 352–367.

(47) Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C., and Womack, T. O. (2017) BUSTER 2.10.3. *Cambridge, United Kingdom Glob. Phasing Ltd.*

(48) Joosten, R. P., Joosten, K., Murshudov, G. N., and Perrakis, A. (2012) PDB_REDO: constructive validation, more than just looking for errors. *Acta Cryst 68*, 484–496.

(49) Karplus, P. A., and Diederichs, K. (2012) Linking Crystallographic Model and Data Quality. *Science (80-.). 336*, 1030–1033.

(50) König, I., Zarrine-Afsar, A., Aznauryan, M., Soranno, A., Wunderlich, B., Dingfelder, F., Stüber, J. C., Plückthun, A., Nettels, D., and Schuler, B. (2015) Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat. Methods 12*, 773–779.

(51) Schuler, B. (2007) Application of single molecule Förster resonance energy transfer to protein folding. *Methods Mol. Biol. 350*, 115–38.

(52) Kudryavtsev, V., Sikor, M., Kalinin, S., Mokranjac, D., Seidel, C. A. M., and Lamb, D. C. (2012) Combining MFD and PIE for Accurate Single-Pair Förster Resonance Energy Transfer Measurements. *ChemPhysChem 13*, 1060–1078.

(53) Best, R. B., Hofmann, H., Nettels, D., and Schuler, B. (2015) Quantitative Interpretation

of FRET Experiments via Molecular Simulation: Force Field and Validation. *Biophys. J. 108*, 2721–2731.

(54) Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model. 51*, 2778–86.

Table 1 - Equilibrium dissociation constants of different peptides binding to Lock 1 and Lock 2.

Peptide	Lock 1 K_D ± std dev (nM)	Lock 2 K_D ± std dev (nM)
KRKRKAK**ITE**	238 ± 4	28 ± 7
KRKRKAK**ITD**	220 ± 20	31 ± 2
KRKRKAK**LTW**	114 ± 2	23 ± 1
KRKRKAK**LSW**	93 ± 6	21 ± 4
KRKRKAK**LSF**	83 ± 10	19 ± 2
KRKAKRK**ASF**	230 ± 54	
KRKAKRK**LAF**	107 ± 34	
KRKAKRK**LSA**	182 ± 38	
KRKRKAK**ITW**	74 ± 2	7 ± 0.3
KRKRKAK**ATW**		30 ± 1
KRKRKAK**IAW**		4 ± 1
KRKRKAK**ITA**		14 ± 4

Table 2 - Equilibrium dissociation constants of the different designs binding the peptide KRKRKAKLSF

Peptide	Lock 1 K_D ± std dev (nM)	AKLSF_design K_D ± std dev (nM)	KLSF_design K_D ± std dev (nM)	LSF_design K_D ± std dev (nM)
KRKRKAK**LSF**	83 ± 10	19 ± 2	51 ± 11	29 ± 2

Table 3 – Kinetic binding constants of the $Y_{III}M_5A_{II}$:$(KR)_5$ complex and the three terminal designed peptide pockets bound to the respective peptides.

Complex	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_D (nM) calculated	K_D ± std dev (nM) from anisotropy
$Y_{III}M_5A_{II}$:$(KR)_5$	$9.7 \cdot 10^8 \pm 1.3 \cdot 10^8$	1.1 ± 0.1	1	1.1 ± 0.8[a]
Lock 1: KRKAKRK**LSF**	$3.3 \cdot 10^8 \pm 0.0 \cdot 10^8$	46.6 ± 3.1	166	83 ± 10
Lock 2: KRKRKAK**LSF**	$3.0 \cdot 10^8 \pm 0.3 \cdot 10^8$	5.5 ± 1.3	20	19 ± 2
Lock 2: KRKRKAK**ITW**	$3.2 \cdot 10^8 \pm 0.0 \cdot 10^8$	1.6 ± 0.0	5	7 ± 0.3

[a] as determined in[15]

Table 4 - Dissociation constants of the respective peptides binding to the internal lock variants

Peptide	Internal Lock 1 K_D ± std dev (nM)	Internal Lock 2 K_D ± std dev (nM)
KRKRKLKFKR	298 ± 45	
KRKRK**A**KFKR	892 ± 47	
KRKRKLK**A**KR	880 ± 28	
KRKAKITWKR		102 ± 19
KRKAK**A**TWKR		733 ± 12
KRKAKI**A**WKR		69 ± 5
KRKAKIT**A**KR		171 ± 20

Figure captions

Figure 1. Peptide binding on dArmRPs (PDB-ID: 5AEI). A Close-up view of a dArmRP with five internal repeats binding a (KR)5 peptide, showing the interactions of the arginine and lysine sidechains (pale green) to the pockets of the dArmRP (yellow-orange). The sidechains are binding to surface pockets and the backbone is kept elongated and fixed by a ladder of conserved asparagines at position 37 (N37). B potential register shifts that can occur due to the repetitive nature of the peptide (red) and the repetitive binding surface (grey, all units are identical and the different shades are only to guide the eye).

Figure 2. Design and structure of Lock 1 in complex with the LSF moiety at 2.1 Å. (A) and (B), Design model with the dArmRP in grey (A) and chain B of the crystal structure in green (B), grafted amino acids in yellow, the LSF peptide moiety in cyan and the 2Fo-Fc-map contoured at 1 σ in black. (C) Schematic interaction map of complex shown in (B) with the coloring according to (B). Only sidechain interactions are shown, hydrogen bonds (length in Å) are displayed in green. The map was created with LigPlot+[54].

Figure 3. Crystal structures of Lock 2 in complex with the AKLSF and AKITW peptide moieties. dArmRPs in green, mutated positions in yellow, peptide in cyan and the 2Fo-Fc-map contoured at 1 σ in black. (A) chain B of Lock 2 with AKLSF bound at 2.1 Å resolution (B) schematic interaction map of complex shown in (A) with the coloring according to B, only sidechain interactions are shown, hydrogen bonds (length in Å) in green. Map created with LigPlot+ (55). (C) Chain D of Lock 2 with AKITW bound at 2.0 Å resolution. (D) Schematic interaction map of the complex shown in (C) with the coloring according to (C), only sidechain interactions are shown, hydrogen bonds (length in Å) in green. Map created with LigPlot+[54].

Figure 4. Register shifts in dArmRP:peptide complexes are observed by single-molecule FRET in solution, and are reduced by incorporating Lock 2. (A) Transfer efficiency histograms of the $Y_{III}M_5A_{II}$:$(KR)_5$ complex (top) and the Lock 2:ITW complex (bottom). The dArmRP was labeled with an acceptor dye (CF660R) at position 51 and the peptide with a donor dye (Cy3B) at the N-terminus. Measured transfer efficiency histograms (gray) are compared to a

fit (red line) with three populations calculated from probability distribution analysis (PDA[36]), assuming that the peak widths are shot noise-limited. The subpopulations share the same peak width and position, and are shaded in green, blue and orange, respectively. Their fitted mean transfer efficiency Efit is given on the top, their relative contributions to both histograms on the right. (B) Cartoon representations of $Y_{III}M_5A_{II}$:KR_5 in the -1 sliding state (top), the optimal bound state (middle), and the +1 sliding state (bottom). The sterically accessible volumes of the donor and acceptor dyes are depicted as green and red clouds. Next to each structure, the calculated mean transfer efficiencies are given.

Figure 5. Crystal structures of the internal Lock pockets with the bound peptides. dArmRPs in green, mutated positions in yellow, peptide in cyan and the 2Fo-Fc-map contoured at 1 σ in black. (A) chain A of internal Lock 1 with the bound peptide, shown is the KLKFKR motif of the peptide (3.17 Å). (B) Chain A of internal Lock 2 with the bound peptide, shown is the AKITW motif of the peptide, the terminal KR residues were not fully resolved in the electron density and only the Cα-atom of the lysine could be fitted (2.8 Å).

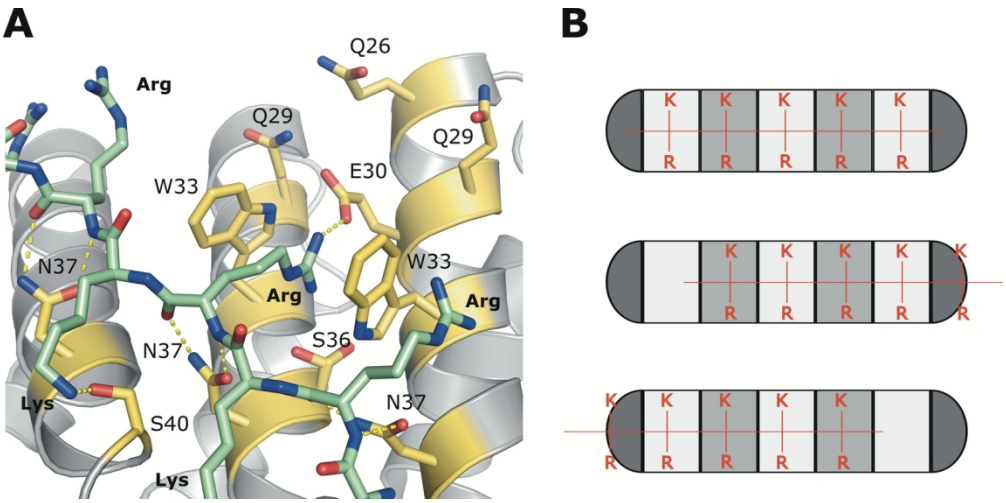

Figure 1. Peptide binding on dArmRPs (PDB-ID: 5AEI). A Close-up view of a dArmRP with five internal repeats binding a (KR)5 peptide, showing the interactions of the arginine and ly-sine sidechains (pale green) to the pockets of the dArmRP (yellow-orange). The sidechains are binding to surface pockets and the backbone is kept elongated and fixed by a ladder of con-served asparagines at position 37 (N37). B potential register shifts that can occur due to the repetitive nature of the peptide (red) and the repetitive binding surface (grey, all units are identi-cal and the different shades are only to guide the eye).

139x68mm (300 x 300 DPI)

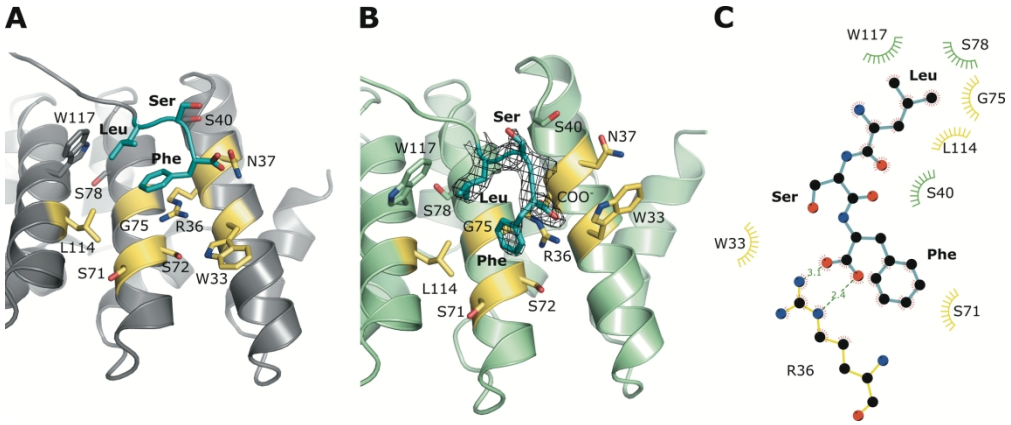

Figure 2. Design and structure of Lock 1 in complex with the LSF moiety at 2.1 Å. (A) and (B), Design model with the dArmRP in grey (A) and chain B of the crystal structure in green (B), grafted amino acids in yellow, the LSF peptide moiety in cyan and the 2Fo-Fc-map con-toured at 1 σ in black. (C) Schematic interaction map of complex shown in (B) with the color-ing according to (B). Only sidechain interactions are shown, hydrogen bonds (length in Å) are displayed in green. The map was created with LigPlot+54.

194x80mm (300 x 300 DPI)

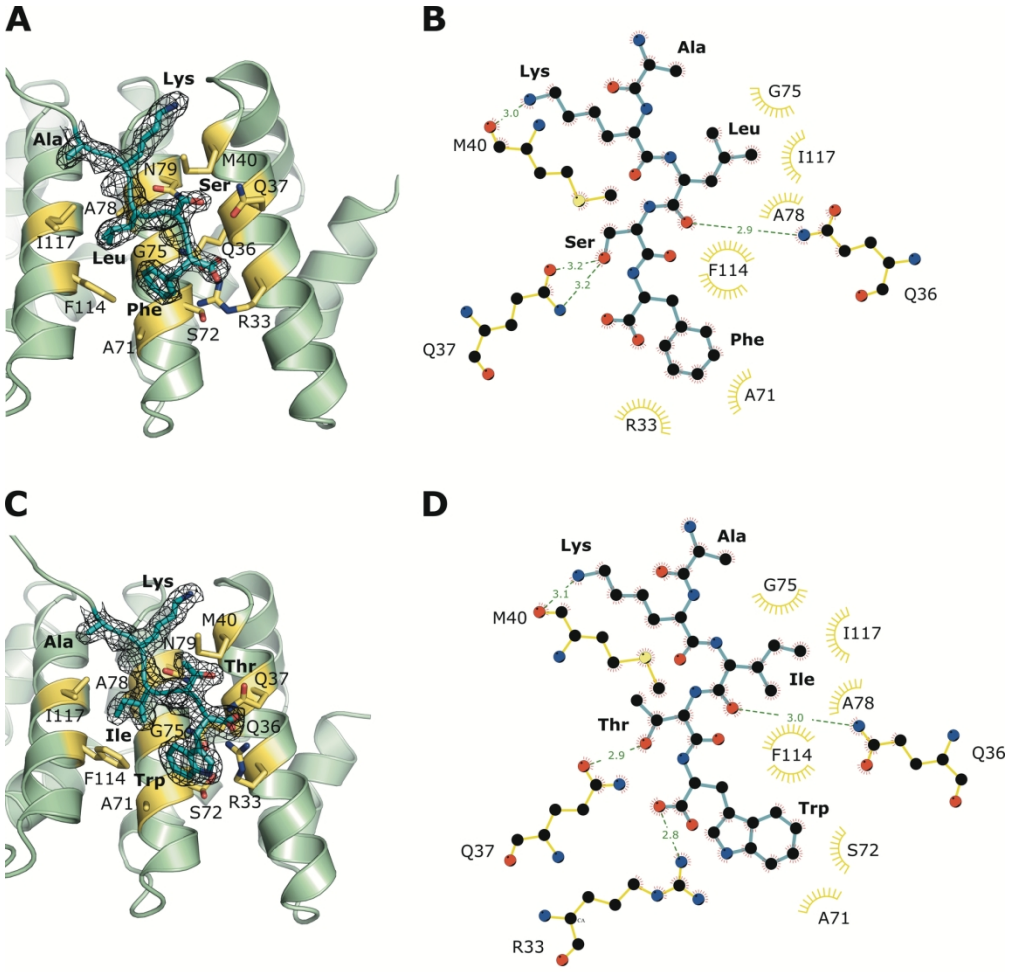

Figure 3. Crystal structures of Lock 2 in complex with the AKLSF and AKITW peptide moie-ties. dArmRPs in green, mutated positions in yellow, peptide in cyan and the 2Fo-Fc-map con-toured at 1 σ in black. (A) chain B of Lock 2 with AKLSF bound at 2.1 Å resolution (B) schematic interaction map of complex shown in (A) with the coloring according to B, only sidechain interactions are shown, hydrogen bonds (length in Å) in green. Map created with LigPlot+ (55). (C) Chain D of Lock 2 with AKITW bound at 2.0 Å resolution. (D) Schematic interaction map of the complex shown in (C) with the coloring according to (C), only sidechain interactions are shown, hydrogen bonds (length in Å) in green. Map created with LigPlot+54.

164x162mm (300 x 300 DPI)

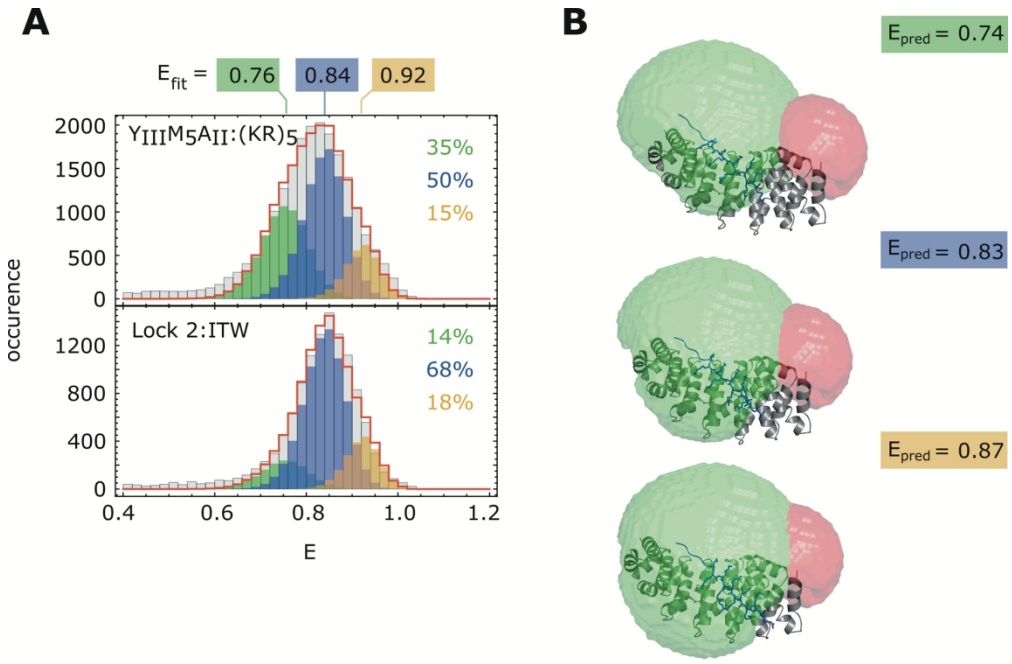

Figure 4. Register shifts in dArmRP:peptide complexes are observed by single-molecule FRET in solution, and are reduced by incorporating Lock 2. (A) Transfer efficiency histograms of the YIIIM5AII:(KR)5 complex (top) and the Lock 2:ITW complex (bottom). The dArmRP was la-beled with an acceptor dye (CF660R) at position 51 and the peptide with a donor dye (Cy3B) at the N-terminus. Measured transfer efficiency histograms (gray) are compared to a fit (red line) with three populations calculated from probability distribution analysis (PDA36), assuming that the peak widths are shot noise-limited. The subpopulations share the same peak width and posi-tion, and are shaded in green, blue and orange, respectively. Their fitted mean transfer efficien-cy Efit is given on the top, their relative contributions to both histograms on the right. (B) Car-toon representations of YIIIM5AII:KR5 in the -1 sliding state (top), the optimal bound state (middle), and the +1 sliding state (bottom). The sterically accessible volumes of the donor and acceptor dyes are depicted as green and red clouds. Next to each structure, the calculated mean transfer efficiencies are given.

154x101mm (300 x 300 DPI)

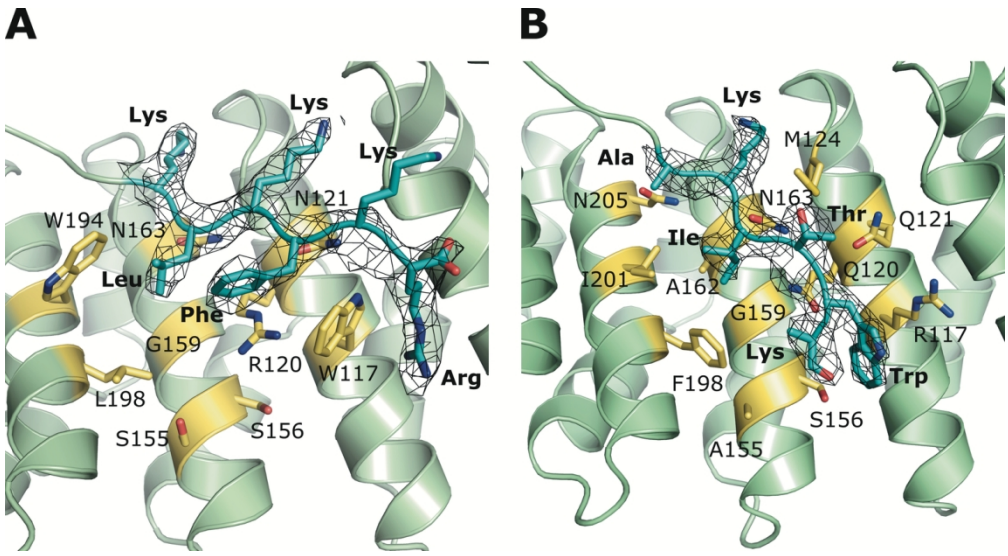

Figure 5. Crystal structures of the internal Lock pockets with the bound peptides. dArmRPs in green, mutated positions in yellow, peptide in cyan and the 2Fo-Fc-map contoured at 1 σ in black. (A) chain A of internal Lock 1 with the bound peptide, shown is the KLKFKR motif of the peptide (3.17 Å). (B) Chain A of internal Lock 2 with the bound peptide, shown is the AKITW motif of the peptide, the terminal KR residues were not fully resolved in the electron density and only the Cα-atom of the lysine could be fitted (2.8 Å).

133x72mm (300 x 300 DPI)

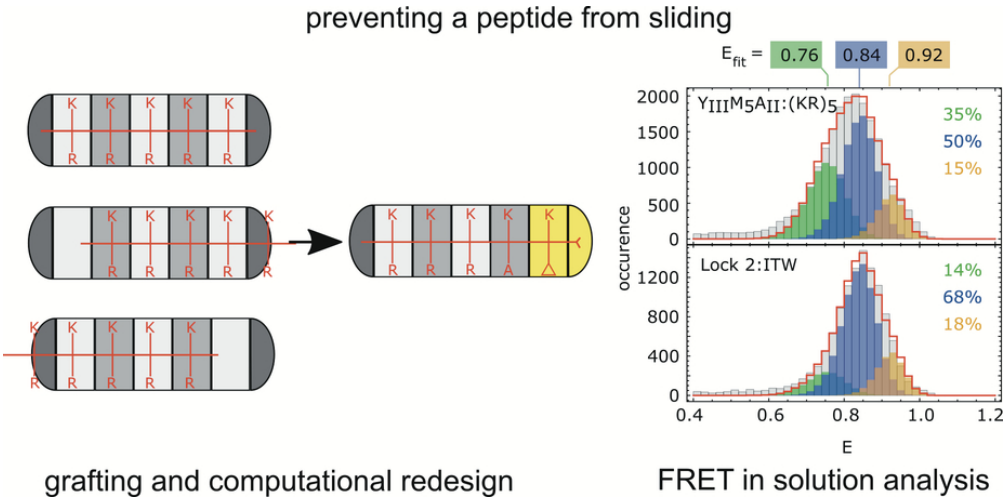

Table of Contents Graphics

78x38mm (300 x 300 DPI)